https://doi.org/10.31449/inf.v49i8.9228

Comparative Performance Analysis of Machine and Deep Learning Models for EEG-Based Biometric Authentication

Ahmad Ayman Tarawneh, Aloui Kamel, and Mohamed Saber Naceur
The University of Sousse by University of Carthage, LTSIRS, INSAT, Tunisa
E-mail: ahmad.tar.tie@gmail.com, kamel.aloui@uvt.tn, medsabeur.naceur@insat.ucar.tn

Keywords: User authentication, physiological signals, electroencephalography (EEG), brain-computer interface (BCI), biometric security

Received: May 14, 2025

EEG-based biometric authentication has emerged as a secure alternative to conventional authentication methods, owing to its resistance to spoofing and inherent movement/image individual variability. This study evaluated the performance of various classification models in the EEG motor movement/image dataset, which comprises 1,526 sessions recorded from 109 subjects using 64 EEG channels at a sampling rate of 160 Hz. A comprehensive set of 1,600 features per session was extracted in the time, frequency, and time-frequency domains. Following standard pre-processing and normalization, the models were trained in a stratified 70/30 training test split using features standardized to zero mean and unit variance.
We systematically compared traditional machine learning classifiers, ensemble methods, and deep learning architectures. Hyperparameter tuning was performed uniformly across all the models. The Ridge Classifier achieved the highest accuracy (93.8%), followed by Logistic Regression (91.27%) and MLP (89.96%), demonstrating the strength of linear and shallow neural models on engineered EEG features. In contrast, deep learning models, including CNN, LSTM, GRU, and BiLSTM, recorded significantly lower accuracy (0.87%) because of limited training data and the use of pre-extracted statistical features instead of raw time-series input, which restricted their ability to learn temporal patterns.
These findings indicate that traditional machine-learning models, when applied to well-crafted features, remain highly competitive for EEG-based authentication. They offer a favorable balance between performance, computational efficiency, and interpretability, whereas deep learning approaches require further adaptation to the structure and scale of EEG data.

Povzetek: Primerjalna študija EEG-biometrije na EEGMMI (109 oseb, 64 kanalov) s 1.600 značilkami pokaže, da linearni modeli prekašajo globoke: Ridge doseže 93,8 % natančnost. Predpripravljene značilke in malo surovih signalov omejita CNN/LSTM; klasični pristopi ostanejo učinkoviti, razlagalni in varčni.

1 Introduction

The increasing number of cybersecurity threats necessitates the implementation of strong user authentication systems to protect sensitive data. Current security frameworks are based on traditional biometric modalities, including fingerprint and iris scans and facial recognition; however, these systems remain exposed to advanced spoofing attacks. High-resolution photographs have been shown to defeat facial recognition systems, whereas synthetic fingerprints made from latent prints have successfully compromised smartphone security [1]. The current limitations of biometric systems demonstrate the requirement for authentication methods that use intrinsic physiological traits that cannot be replicated.

EEG is a promising noninvasive brain activity recording method that shows potential as a biometric solution. EEG signals produce dynamic neural patterns that function as individual-specific brain fingerprints because they differ from the static physical characteristics [2]. The nature of EEG signals binds them to life sciences; users must actively participate in recording and playback. Research shows that EEG responses to visual flashes, auditory tones, and motor imagery tasks produce different patterns between subjects, which allows for effective user identification [3].

Despite their potential, EEG-based verification faces significant challenges. Signal variability caused by environmental noise, electrode displacement, or shifts in user mental states (e.g., fatigue and stress) can degrade the performance over time [4]. In addition, most existing systems rely on high-density electrode arrays (e.g., 64–128 channels), which are impractical for everyday use in consumer devices.

The integration of physiological signals into authentication systems has transformative implications across various industries. In personal devices, continuous authentication using EEG or photoplethysmography (PPG) can enable seamless yet secure access to smartphones and wearables, thereby reducing reliance on vulnerable password-based systems [5]. In addition, studies suggest that EEG-driven authentication can enhance financial transactions by adding a robust layer of identity verification, mitigating fraud risks, and improving digital security [6] [7].In healthcare, physiological biometrics can safeguard electronic health records (EHRs) and restrict access to sensitive medical devices, aligning with regulatory mandates such as HIPAA and GDPR [8]. In the changing landscape of security challenges, there are opportunities to improve authentication systems by combining neuroscientific insights and biometric technology with physiological signals [9].

Standard text-based passwords are commonly used. Most users tend to opt for passwords or use them repeatedly across platforms, which increases their risk of entry. Furthermore, cryptic passwords can be challenging to recall, resulting in users resorting to less-secure alternatives [8]. Research indicates that although graphical passwords and session-based authentication enhance security to some degree, they remain susceptible to attacks and usability issues [9] [10]. Proposals have been made to use biometrics, such as keystroke dynamics, to tackle these vulnerabilities without the need for hardware. However, these methods also encounter difficulties in terms of accuracy and environmental reliance [11].

1.1 Research objectives and questions

This study aims to advance EEG-based biometric verification by evaluating the effectiveness of spectral and temporal features extracted from EEG signals. We utilized a publicly available dataset containing recordings from 64 EEG channels and assessed model performance across multiple sessions. A variety of machine learning and deep learning classifiers were applied to determine their accuracy and reliability in user authentication. Our work contributes to the broader goal of developing secure and practical EEG-based biometric systems by providing a comparative performance analysis of commonly used classification models.

The primary objective of this study is to evaluate the effectiveness of various machine learning and deep learning models in biometric authentication based on EEG. Specifically, we investigate the following:

- RQ1: Can low-complexity machine learning models, such as Ridge Classifier and Logistic Regression, achieve high accuracy in EEG-based biometric authentication?

- RQ2: How do deep learning models perform compared to traditional machine learning models when applied to extracted statistical features from EEG data?

- RQ3: Under what conditions (e.g., data volume, feature types) could deep learning models outperform traditional machine learning models in EEG-based authentication tasks?

By addressing these questions, we aimed to provide insights into the suitability of different classification approaches for EEG-based biometric systems.

1.2 Organization of the paper

The remainder of this paper is organized as follows.

Section 2 provides background information on the EEG fundamentals, electrode placement, frequency bands, and the general framework for EEG-based authentication. Section 3 reviews related work in the field of EEG-based biometric authentication. Section 4 describes the datasets used in this study. Section 5: Details of the experimental setup including data filtering, outlier analysis, feature extraction, normalization, dataset splitting, performance metrics, and classification models. Section 6 presents the results of the classification models. Section 7 discusses the findings, compares the performance of different models, and analyzes the conditions that influence their effectiveness. Section 8: Concludes this study and suggests directions for future research.

2 Background

The EEG-based authentication leverages the unique neural activity of the brain to create a robust and secure biometric system. Unlike traditional authentication methods, EEG signals are inherently tied to an individual's cognitive and physiological state, making them difficult to replicate or forge. This section explores the fundamentals of EEG, its signal frequency bands, and the optimal electrode placement for enhancing biometric accuracy.

2.1 Fundamentals of EEG and its role in authentication

This study focuses on EEG-based biometric authentication using machine and deep learning models. Electroencephalography (EEG) is a widely used physiological signal in Brain–Computer Interface (BCI) research due to its ability to capture unique brainwave patterns that are difficult to replicate [12, 13]. Although BCI systems often aim to integrate multiple modalities and interactive capabilities, the scope of this work is limited to the use of EEG signals alone for identity verification. The motivation stems from BCI principles, but the objective here is not to develop a full BCI framework, rather to assess the effectiveness of EEG-based classification models for secure user authentication.

EEG functions as a method to detect brain electrical signals that combine the synaptic potential activity of multiple cerebral cortex neurons [14]. This technique enables the simultaneous multichannel measurement of central and autonomic nervous system responses. The central nervous system, which consists of the brain and spinal cord, reacts to external stimuli, and the autonomic nervous system controls involuntary body processes, including heart rate and breathing [15]. EEG signals serve as reliable measures of neural activity triggered by both internal and external stimuli and reflect unique physiological and behavioral traits.

One of the key advantages of EEG for authentication is its uniqueness and difficulty in replication. EEG signals contain individualized characteristics, such as cognitive ability, emotional state, age, gender, and neural connectivity [16, 17, 18]. Because brain structures and cognitive functions vary among individuals, EEG signals exhibit substantial inter-subject differences but remain stable when the same individual performs identical tasks [19, 20]. Furthermore, EEG authentication is highly resistant to spoofing because it requires specialized recording equipment, unlike facial or fingerprint recognition, which can be easily compromised [21].

Standard text-based passwords remain the most widely used authentication mechanism; however, their vulnerabilities are well documented. Users frequently reuse simple passwords across multiple platforms, increasing the risk of credential theft [8], whereas complex passwords are often abandoned because of memorability challenges [22]. Although graphical and session-based alternatives mitigate some risks, they remain susceptible to shoulder surfing, brute-force, and replay attacks [9, 10]. Behavioral biometrics, such as keystroke dynamics, offer hardware-free solutions but struggle with accuracy under variable user states (e.g., fatigue) or environments [11]. These shortcomings underscore the need for systems that balance the security, usability, and robustness. Physiological signals, such as EEG, bypass these issues by exploiting intrinsic biological traits that are resistant to spoofing and memorization [4].

These shortcomings underscore the need for authentication systems that balance the security, usability, and robustness. Physiological signals, such as EEG, offer a promising alternative by exploiting intrinsic biological traits that are resistant to spoofing and are independent of user memorization [4].

2.2 EEG electrode placement

EEG authentication accuracy is significantly influenced by electrode placement because different brain regions generate distinct responses to cognitive and sensory stimuli [23]. The selection of appropriate electrode positions plays a critical role in improving the recognition rate and reducing the complexity of the data collection. Figure 1 shows the placement of the 64-channel EEG sensors used in the BCI2000 system to capture motor and imagery task-related brain activity.

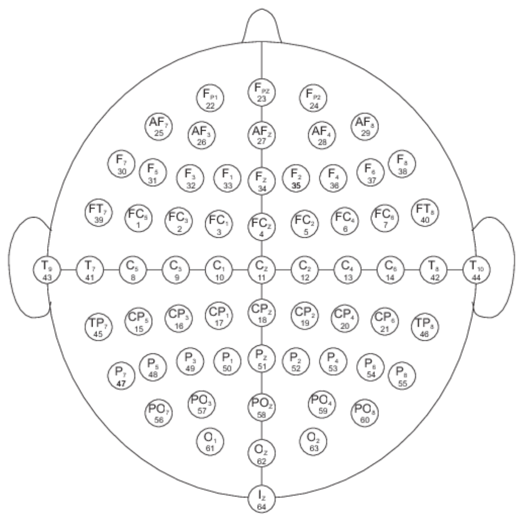

Figure 1: 64-channel EEG electrode distribution [24]

Several studies have identified optimal electrode locations for EEG-based biometric authentication. [25] found that the O2 channel provides stable biometric features in semantic-induced ERP paradigms . [26] highlighted key authentication features in the Fz, FC1, FC2, Cz, CP1, CP2, and Pz channels, while [27] emphasized PO3, PO4, O1, Oz, and O2 as effective biometric authentication regions. These findings suggest that authentication accuracy can be maximized by strategically selecting electrode placements based on the task-specific requirements.

Because EEG patterns vary among individuals, selecting a personalized set of EEG channels can further enhance the authentication performance [28]. [29] proposed an optimized channel-based model that dynamically adjusts electrode locations per user, thereby improving robustness and reducing the overall data collection burden. In addition, researchers have explored the use of genetic algorithms to refine authentication channel selection, demonstrating further improvements in EEG-based biometric accuracy [28].

This study demonstrates how EEG signal optimization, frequency band selection, and electrode placement affect authentication systems. Researchers continue to improve EEG biometric systems by fine-tuning these factors, resulting in more secure and reliable systems that can be used in real-world applications.

2.3 EEG signal frequency bands

EEG signals consist of separate frequency bands that correspond to the different neural states of activity. The selection of appropriate frequency bands for authentication systems improves accuracy while reducing computational requirements [30]. Researchers have grouped EEG signals into multiple frequency bands with unique characteristics for biometric security applications.

The delta band shows the most distinguishable characteristics, making it suitable for identity recognition because it remains consistent between different states [30]. Beta and gamma bands achieve better authentication accuracy because they correlate with cognitive and visual-related mental tasks [28, 20, 31]. The gamma band exhibits strong authentication potential because its chaotic and complex nature leads to strong nonlinearity [32].

Identification of identity-related EEG features requires more than one frequency band because no single band contains all necessary information. The spread of biometric information across various frequency bands requires an integrated method using multiple frequency components [33, 34, 7]. Authentication accuracy varies because stimulation tasks produce EEG responses that differ across specific frequency bands [28]. Authentication frameworks must adapt their performance to specific EEG responses from different tasks to achieve optimal results.

Table 1: EEG frequency bands and their characteristics

Freq. Band	Range (Hz)	Typical Amplitude	Dominant Brain Region
Delta	from 1 to 4 Hz	from 20 to 200 µV	Frontal and occipital lobes
Theta	from 4 to 8 Hz	from 100 to 150 µV	Frontal and parietal lobes
Alpha	from 8 to 13 Hz	from 20 to 100 µV	Parietal lobes and posterior occipital
Beta	from 13 to 30 Hz	from 5 to 20 µV	Central areas, temporal and frontal lobes
Gamma	greater than 30Hz	less than 2 µV	Somatosensory center

Table 1 shows the essential characteristics of the EEG frequency bands, including their frequency range and amplitude, together with their main brain regions. The different cognitive and physiological states of EEG-based authentication systems depend on the frequency bands. The unique features of each band allow researchers to enhance biometric accuracy through optimized feature extraction and classification methods.

2.4 General framework

The general framework for EEG-based person identification systems appears in Figure 2. Identification systems based on EEG data follow a specific operational sequence that includes multiple essential phases. EEG signals are acquired through scalp electrodes.

The raw signals receive preprocessing treatment, which includes noise reduction, artifact removal, and normalization steps to improve the signal quality. This system uses spectral, temporal, or spatial analysis techniques for feature extraction to detect specific neural patterns. The extracted features are classified into classification models, which include machine learning and deep learning algorithms, to distinguish people using their brainwave signatures. The system uses the classifier output to perform identity verification or authentication while maintaining a secure and reliable identification process.

3 Related works

Physiological signals such as electroencephalography (EEG), electrocardiography (ECG), and heart rate variability (HRV) offer promising alternatives to traditional authentication methods. Unlike static biometrics (e.g., fingerprints), which are vulnerable to spoofing through synthetic replication, physiological signals are intrinsically tied to dynamic biological processes. Among these, EEG has emerged as the gold standard because of its neurophysiological uniqueness and inherent live-ness detection capabilities.

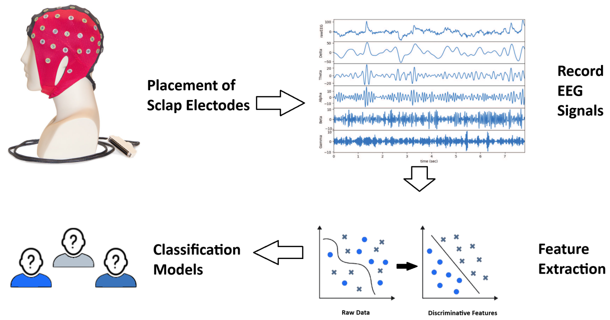

Figure 2: General framework of EEG-based person identification systems

Recent advancements have refined EEG-based authentication through innovations in sensor technologies and signal processing. For example, research on brain-machine interfaces (BMIs) has provided deeper insights into mental state classification using EEG signals, reinforcing the viability of this biometric approach [45]. Furthermore, [46] provided a comprehensive review of sensor modalities for brain-computer interfaces, emphasizing the strengths and limitations of EEG technology in authentication applications, and hybrid EEG-MEG systems, proposed by [47], address signal quality limitations by combining EEG temporal resolution with MEG spatial precision, although practical deployment remains constrained by hardware complexity [35]. Recent advances in EEG/MEG source imaging have improved signal quality and enhanced authentication reliability [48].

Electrode optimization is a critical focus for practical EEG systems portability of EEG authentication as demonstrated by [36], who proposed a blink-induced EEG system for mobile devices. Using a 14-channel Emotive EPOC headset, EEG signals were recorded during natural eyeblinks from 30 participants. A support vector machine (SVM) classifier achieved 92% accuracy by analyzing delta-band (0.5–4 Hz) power changes associated with blink-related neural activity. This study highlighted EEG's potential of EEG for zero-effort authentication in mobile contexts, although electrode density and user comfort remain barriers. Similarly, [35] proposed genetic algorithms to dynamically optimize electrode placement per user, thereby reducing intersession variability. These efforts highlight the tradeoff between usability (fewer electrodes) and robustness, which is a central challenge in EEG biometrics.

Recent advancements in EEG-based biometric authentication have yielded promising results. [41] employed an eight-channel OpenBCI headset to collect EEG data from

Table 2: Summary of EEG-based biometric authentication studies

Study	EEG Task	No. of Electrodes	Classifier	Accuracy
[3]	User Identification (9 Subjects)	64	GMM with MAP Adaptation	Best Half Total Error Rate 7.7
[35]	Four Mental Imagery Task	64	Genetic Algorithm + SVM	97.69% - 100%
[36]	Authentication for Mobile Devices	14	SVM	92%
[37]	Continuous Authentication	4	Hybrid LSTM-CNN	EER: 1.8%
[38]	Image Classification via EEG	Not Specified	CNN	70% (EEG Only), 82% (EEG + Image Features)
[39]	Modeling Biosignals	Not Specified	Contrastive Learning	81.6% and 93.2%
[40]	Person identification	Not Specified	Autoencoder-CNN	87.6%
[41]	Person Classification	8	SVM	92.9%
[42]	Event-Related Potential (ERP)	Not Specified	CNN + GCNN (EEG-BBNet)	99.26%
[43]	Epileptic seizure detection (healthy vs epileptic and ictal vs seizure-free)	21	k-NN, SVM, ANN	Up to 99% with DWT features
[44]	Confusion States	1 (commercial EEG headset)	1D CNN	99%

12 subjects during fatigue analysis tasks. They extracted ten features per channel and utilized a multiclass Support Vector Machine (SVM) classifier, achieving a maximum identification accuracy of 92.9% using a radial basis function kernel. This study highlights the potential of EEG signals for reliable user authentication.

In another study, [42] introduced EEG-BBNet, a hybrid framework combining Convolutional Neural Networks (CNN) with Graph Convolutional Neural Networks (GCNN) to capture both spatial and connectivity features of EEG signals. Evaluated on a benchmark dataset encompassing various brain-computer interface tasks, EEG-BBNet achieved an average correct recognition rate of up to 99.26% in event-related potential tasks using intrasession data. The model demonstrated robustness across different connectivity measures and maintained a high performance even with a reduced number of electrodes, highlighting its practicality for real-world applications.

In the summary table (Table 2), these studies illustrate the efficacy of advanced machine-learning techniques and hybrid neural network architectures in enhancing EEG-based biometric authentication systems.

Recent advances in EEG-based classification tasks have demonstrated the effectiveness of both traditional feature extraction and deep-learning approaches in various cognitive and clinical contexts. [54] investigated epilepsy detection using EEG signals by comparing three feature extraction techniques: time-domain statistical features, frequency-domain features via Discrete Cosine Transform (DCT), and time-frequency features using Discrete Wavelet Transform (DWT). Their experiments, conducted on the Bonn EEG dataset, showed that DWT-based features yielded the highest classification performance, particularly when distinguishing between ictal and seizure-free states, achieving accuracy levels comparable to or exceeding those of the existing state-of-the-art methods. In a different application domain, [44] explored the detection of confusion in students during video lectures by using EEG recordings. They extracted features across multiple EEG frequency bands and trained a one-dimensional Convolutional Neural Network (1D-CNN) to classify confusion states. Although the specific accuracy figures were not disclosed, the proposed deep learning model significantly outperformed traditional machine learning approaches, highlighting the potential of EEG-driven models for real-time cognitive state monitoring in educational environments. Complementing these EEG studies, [43] focused on motor imagery detection using ECG signals derived from the PhysioNet EEG Motor Movement/Imagery dataset. Their model employed Wavelet Packet Decomposition for multiresolution feature extraction and a multiscale convolutional neural network (MSCNN) for classification. The system achieved strong performance metrics (92% accuracy, 91% F1-score, and 95% ROC-AUC), underscoring the potential of combining advanced signal decomposition with deep learning architectures for decoding motor intentions, a technique that may also be adapted to EEG-based BCI applications.

The foundational work of [3] established EEG's viability of EEG as a biometric identifier. Using maximum a posteriori (MAP) model adaptation, Marcel and Millán demonstrated that EEG responses to visual and motor imagery tasks could achieve 95% user identification accuracy across a cohort of nine subjects. Their methodology involved extracting spectral features (alpha and beta bands) from 64-

Table 3: Comparison of recent studies on physiological signal-based authentication: key tasks, and electrode positions

REF	Tasks	No Of Electrodes	Positions
[7]	MI (Motor Imagery)	19	O2, O1, P8, P7, P4, Pz, P3, C4, Cz, C3, T8, T7, F8, F7, Fz, F4, F3, Fp2, Fp1
[49]	MI	17	O2, O1, T6, T5, P4, P3, PZ, T4, T3, C4, C3, CZ, F8, F7, F4, F3, FZ
[29]	ERP (Event-Related Potential)	16	Cp6, Cp5, Af8, Af7, F4, F3, C4, C3, Po8, Oz, Po7, P4, Pz, P3, Cz, Fz
[50]	ERP	14	O2, O1, T8, T7, P8, P7, FC6, FC5, F8, F7, F4, F3, AF4, AF3
[33]	VEP (Visual Evoked Potential)	14	O2, O1, T8, T7, P8, P7, FC6, FC5, F8, F7, F4, F3, AF4, AF3
[51]	VEP + sound	14	O2, O1, P8, P7, T8, T7, FC6, FC5, F8, F7, F4, F3, AF4, AF3
[52]	Resting state	6	O2, O1, P8, P7, C4, C3
[53]	VEP	6	Oz, O2, O1, Pz, Cz, Fpz

channel EEG data and employing Gaussian mixture models (GMMs) for classification. A key innovation was the use of MAP adaptation to personalize generic models to individual users, thereby reducing intersession variability. However, the reliance on high-density electrode arrays limits their practical deployment.

To address real-world usability, [37] introduced a continuous authentication system using a 4-channel EEG headset. Their hybrid LSTM-CNN architecture processed theta (4–8 Hz) and gamma (30–50 Hz) band features, achieving a 1.8% equal error rate (EER) over 12 sessions with 50 subjects. The strength of the system lies in its resilience to short-term signal variability (e.g., mood changes), although performance degraded by 4% in noisy environments. This study emphasized the trade-off between usability (fewer electrodes) and robustness, which is a challenge central to EEG biometrics.

Deep learning models efficiently process vast amounts of visual data; however, their decision-making process remains opaque. Recent research [38] introduced methods to extract image features from EEG signals, thereby enhancing model interpretability and convergence efficiency. Inspired by this, EEG signals were encoded as images to improve the brain signal analysis using deep learning. By classifying EEG representations corresponding to 39 image classes, researchers achieved a benchmark accuracy of 70%, surpassing previous methods. Furthermore, integrating EEG-based features with conventional image classifiers resulted in 82% accuracy, thereby demonstrating the potential of EEG-enhanced deep learning for improved classification and biometric applications.

Researchers [39] who employed a self-supervised contrastive learning approach demonstrated promising results in EEG-based classification, particularly in handling inter-subject variability and noisy labels. Using the same EEG Motor Movement/Imagery Dataset (EEGMMI), the study achieved a recognition accuracy of 88.6%, highlighting the effectiveness of contrastive learning in modeling EEG signals with a reduced reliance on labeled data. The research enhanced representation quality through subject-aware learning techniques, which included subject-specific contrastive loss and adversarial training, to achieve competitive classification results compared to fully supervised methods.

A recent study [40] applied the same EEG Motor Movement/Imagery Dataset (EEGMMI) to demonstrate that deep learning models work well for EEG-based person identification. The research used an autoencoder-CNN framework to achieve 87.60% recognition accuracy for task-based identification, which demonstrates the ability of deep learning to identify people through their EEG signals.

Table 3 summarizes various EEG-based authentication studies, including the type of tasks performed, number of participants, and electrode positions used in each study. The referenced studies covered a range of tasks such as Motor Imagery (MI), Event-Related Potentials (ERP), and Visual Evoked Potentials (VEP), among others, with specific electrode placements listed for each study.

4 Methodology

This study aims to develop and evaluate EEG-based biometric authentication models by leveraging a well-structured experimental pipeline comprising data acquisition, preprocessing, feature engineering, model training, and statistical evaluation. The overall methodology was structured to ensure reproducibility, generalizability, and rigorous assessment of the classifier performance. Figure 3 illustrates the key stages of this methodology.

4.1 Data acquisition and preprocessing

The EEG Motor Movement/Imagery (EEGMMI) dataset was selected for its comprehensiveness and suitability for

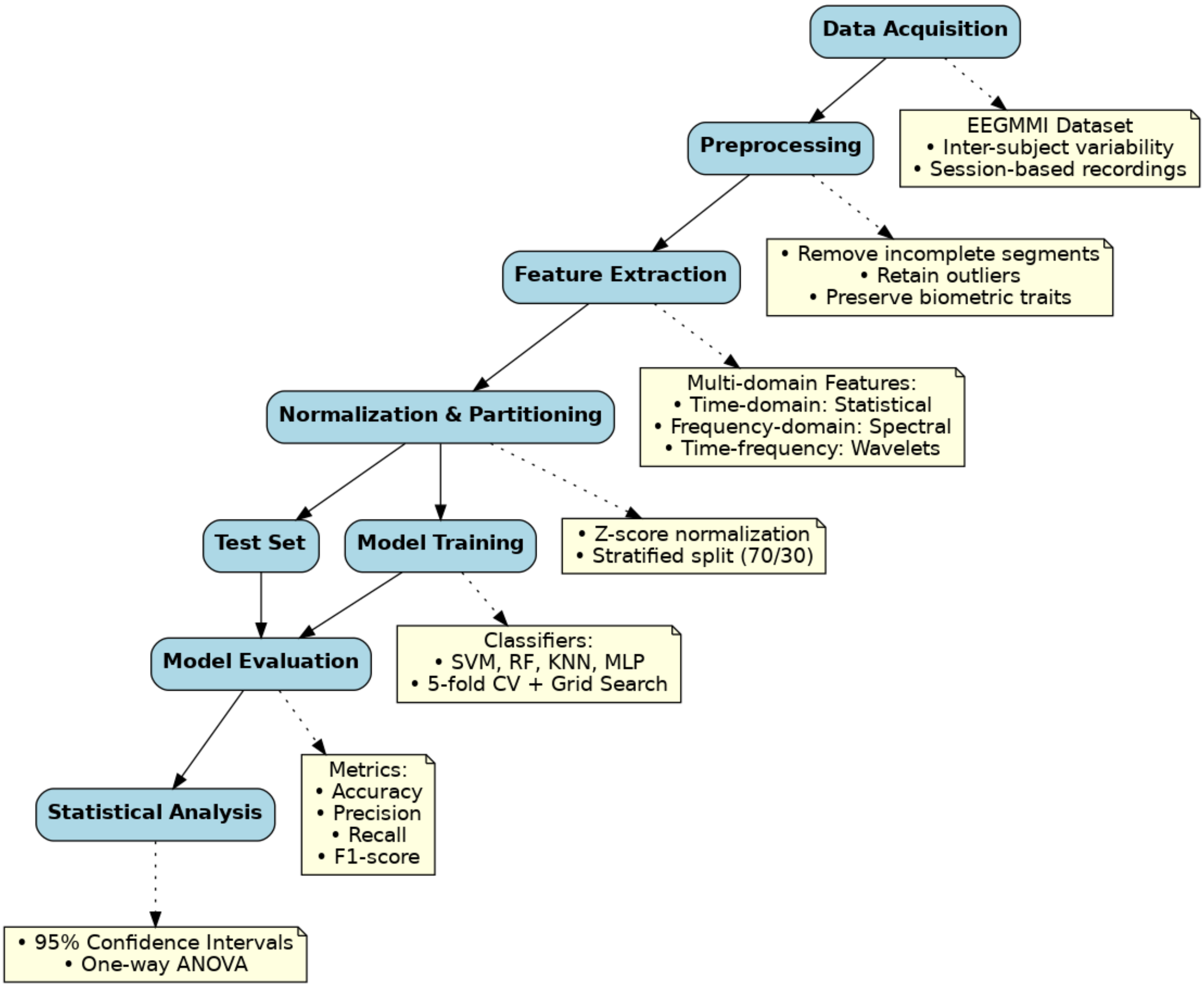

Figure 3: Overview of the methodology pipeline for EEG-based authentication

biometric research. It offers high intersubject variability and session-based recordings, which are crucial for evaluating the consistency of neural signatures over time. Preprocessing focuses on data cleaning to remove incomplete signal segments while preserving valid physiological patterns. Outliers were visually analyzed but retained to preserve the integrity of individual biometric traits, consistent with prior biometric research practices [3].

4.2 Feature extraction strategy

To capture the complexity of the brain signals, features were derived from the time, frequency, and time-frequency domains. This multidomain approach increases the discriminative power of EEG data by capturing both static and dynamic signal properties. The time-domain features highlight statistical properties, the frequency domain captures oscillatory behavior via spectral power analysis, and the time-frequency domain enables the detection of transient and nonstationary events using wavelet decomposition [55]. The feature set was deliberately designed to maintain interpretability and robustness, while ensuring sufficient dimensionality for machine learning models.

4.3 Normalization and data partitioning

All features were standardized using z-score normalization to ensure fair comparisons between features measured on different scales, which is a critical step for algorithms sensitive to distance metrics, such as SVM and KNN [56]. Stratified train-test splitting was performed to preserve the class distributions in both subsets, allocating 70% of the data for training and 30% for testing. This division allows the model generalization to be evaluated using unseen data.

4.4 Model selection and evaluation protocol

To benchmark the effectiveness of the EEG-based biometric authentication, we selected a broad set of classifiers from different algorithmic families: linear models, tree-based ensembles, probabilistic classifiers, distance-based methods, and deep learning architectures. Each classifier was subjected to stratified 5-fold cross-validation on the training set to ensure robust performance estimation and minimize overfitting. The hyperparameters were optimized uniformly across the classifiers using a grid search.

The evaluation was based on four widely accepted classification metrics: accuracy, precision, recall, and F1-score. These metrics collectively provide insight into model correctness, sensitivity to false positives and false negatives, and a balance in handling class distributions. The final model evaluation was conducted on the held-out test set, and the performance was further analyzed using 95% confidence intervals and one-way ANOVA tests to determine the statistical significance of differences across classifiers.

4.5 Statistical rigor and reliability

To ensure the statistical reliability of the results, we conducted a confidence interval estimation using the Student's t-distribution because of the limited number of folds ($n = 5$). One-way ANOVA tests were applied to compare the classifier means across each performance metric, providing statistical evidence of significant differences. This analytical rigor ensures that the observed performance gaps are not due to random variance, thus supporting reliable conclusions regarding model performance.

5 Dataset

This study uses the EEG Motor Movement/Imagery Dataset (EEGMMI) from PhysioNet [57], which serves as a widely recognized public database for EEG research. The dataset was chosen because it contained numerous subjects alongside multiple recording sessions for each participant, thus making it ideal for EEG authentication system development and testing.

The dataset contains 109 subjects, which provides researchers with a diverse population to study. A large dataset size provides strong authentication model reliability because it covers various neural patterns across different individuals. The dataset contained 14 recording sessions for each subject, which enabled researchers to measure the session-to-session variability. Multiple recording sessions enable researchers to test the time-dependent stability of EEG-based biometric features, which is essential for developing dependable authentication systems.

The EEGMMI dataset benefits from its well-defined experimental design that combines motor execution with motor imagery tasks. These tasks produce separate neural responses that serve as an effective base for extracting features and conducting classifications. The EEG system uses 64 channels to record data while following the 10-10 electrode placement system, which is recognized worldwide. The high-density setup provides precise spatial resolution of brain activity, which allows for a better analysis of neural patterns that are important for biometric authentication.

The dataset benefits from its 160 Hz sampling rate, which provides sufficient resolution to detect neural oscillations. The two-minute recording sessions delivered sufficient data for analysis through a practical balance between data volume and usability. The dataset also available on Kaggle: https://www.kaggle.com/datasets/brianleung2020/eeg-motor-movementimagery-dataset

6 Experimental setup

The experimental setup of this study involved a structured pipeline for processing and analyzing EEG signals for biometric authentication. The process starts with data preprocessing, in which missing values and outliers are checked to ensure data quality. Feature extraction is then performed to extract relevant statistical, spectral, and time-frequency features from EEG signals. This section explains the steps taken to improve the dataset, remove outliers, and extract features that are useful for the classification and authentication of individuals.

6.1 Data filtering

To ensure data quality, preprocessing was performed to address missing values. EEG signals recorded across 64 channels sometimes have null or zero values because of recording artifacts or hardware limitations. Instead of discarding all incomplete rows, only rows in which all 64 channels contained zero or null values were removed.

This decision was made to retain meaningful EEG data because zero values in specific channels can represent valid physiological states rather than missing data.

Figure 4 illustrates the distribution of the null values across the dataset. In total, 56,548 rows were entirely null across all channels and discarded. Each of the 109 subjects participated in 14 sessions, with an average of 18,226 rows per session, resulting in a final dataset comprising 27,756,828 rows of EEG signals spanning all sessions and subjects.

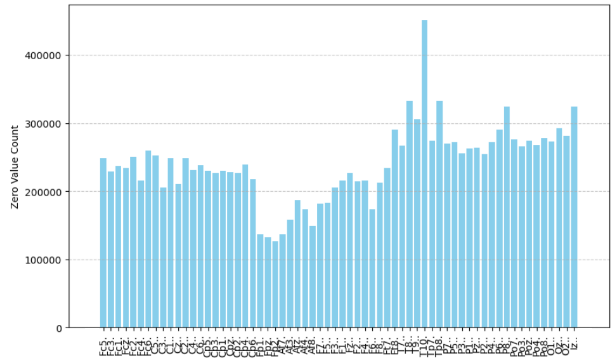

Figure 4: Bar chart displaying the count of zero values in each column of the dataset

6.2 Outlier analysis

Outlier detection was conducted to examine the presence of extreme values in the EEG signals, which could result from physiological variations, noise, or sensor artifacts. However, because the primary objective of this study is user verification, modifying or removing these outliers could distort the individual neural signatures and negatively affect authentication accuracy.

To visualize outliers more clearly, a data reduction approach was applied: the EEG signals were grouped into non-overlapping windows of 1,000 samples, and the mean value for each window was computed. This process reduced visual noise while preserving the general distribution characteristics per channel. Figure 5 presents boxplots of the aggregated values for each EEG channel, providing an overview of data dispersion and outlier presence. Although extreme values are visible, they were retained to maintain the integrity of subject-specific EEG patterns for biometric analysis.

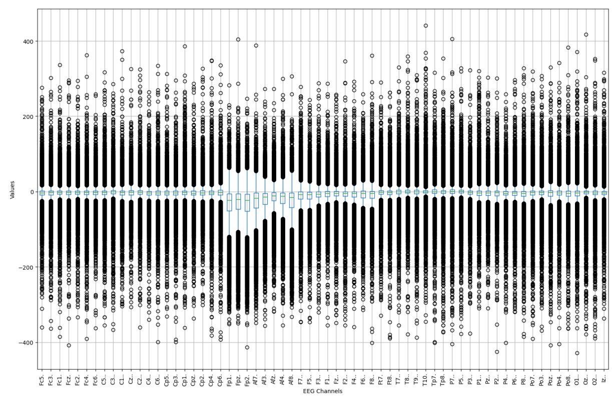

Figure 5: Boxplot visualization of reduced EEG signal distributions across individual channels.

6.3 Feature extraction

The original dataset consisted of 14 files per subject, corresponding to 109 subjects, resulting in 1,526 files. Each file contained EEG recordings from 64 channels over durations ranging from one to two minutes. Following preprocessing, feature extraction was conducted across three domains: frequency, time, and time-frequency. The **time-domain** features were used to describe the statistical properties of the EEG signals, including the minimum, maximum, mean, standard deviation, variance, kurtosis, and skewness. These features are important for amplitude fluctuations of the signal and help in understanding individual neural activity.

In the **frequency domain** , spectral power features were extracted using the Short-Time Fourier Transform (STFT) to analyze EEG activity across different frequency bands. These bands include gamma (30–50 Hz), beta (13–30 Hz), alpha (8–13 Hz), theta (4–8 Hz), and delta (0.5–4 Hz) bands, each associated with distinct neural processes and cognitive states. The power distribution across these frequency bands serves as a critical factor in biometric identification because variations in spectral content provide a unique signature for each individual.

For the **time-frequency-domain** analysis, a wavelet transform was employed to capture transient and non-stationary EEG characteristics. Wavelet-based features include band energies and entropies derived from multiple decomposition levels, allowing for multi-resolution analysis of EEG signals. The extracted wavelet features provided additional information on the time- and frequency-domain features, thus improving the overall discriminative power of the dataset.

Because EEG data are multidimensional, a total of 25 features were extracted from each of the 64 channels, resulting in 1,600 features per session 25×64. With 1,526 EEG session files, the final dataset was structured as a feature matrix of size $1,526 \times 1,600 = 2,441,600$, where each row represents a session and each column represents a specific extracted feature. This structure ensures a comprehensive representation of EEG signals for biometric authentication.

6.4 Feature normalization and dataset splitting

Normalization was performed on the selected EEG features after removing the missing values and outliers to ensure uniformity in the data distribution. The feature values were standardized to have zero mean and unit variance for all the features. It is essential to improve the performance of the classification model, especially for the k-nearest neighbors (KNN) and support vector machine (SVM) algorithms that use distance measures. Normalization also helps reduce the effect of different scales in the dataset, thus enabling classifiers to learn better.

Next, the dataset was split into training and testing sets for use in model evaluation. Stratified sampling was used to preserve the proportion of each class, thus providing a balanced representation of both the training and test sets. The dataset was split into 70% for training the classification models, and 30% for testing. This split helps avoid overfitting and enables trained models to generalize well to unseen data.

6.5 Performance metrics

The classification model evaluation was performed using four main performance metrics: **accuracy, precision, recall, and F1-score**. Accuracy is a measure of the total number of correct predictions in comparison to the total number of predictions made by the model. Precision measures the proportion of true positives to all positive predictions, as it is crucial in scenarios that need to minimize false positives. Recall measures a model's ability to correctly identify positive instances by comparing it to the total number of actual positive cases. The F1-score is a balanced performance metric, which is the harmonic mean of the precision and recall to address class imbalance. These metrics enable a holistic assessment of the classification performance, which provides information about each model's advantages and disadvantages.

7 Statistical analysis and results

To ensure the effectiveness of EEG-based biometric authentication, different classification models from various algorithm families, including traditional machine learning, ensemble learning, and deep learning approaches, were utilized. The chosen models included Random Forest, Ridge Classifier, Logistic Regression, Calibrated Classifier CV, XGBoost, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), histogram boosting, Gaussian naïve Bayes (GaussianNB), multilayer perceptron (MLP), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Bidirectional LSTM, and Convolutional Neural Networks (CNN).Hyperparameter tuning was performed uniformly for all the classifiers to enhance their performance.

In this section, we present a comprehensive statistical validation of the 12 classifiers evaluated using our EEG-based dataset. We begin by detailing the cross-validation procedure and reporting fold-wise results for accuracy, precision, recall, and F1-Score. Next, we computed 95% confidence intervals (CIs) for each metric. We then performed one-way analysis of variance (ANOVA) tests to assess whether the observed differences among the classifiers were statistically significant. Finally, we provide the performance metrics for a holding test set. We refer to the corresponding figures and tables in the text.

7.1 Cross-validation metrics

We performed a five-fold stratified cross-validation for each of the 12 classifiers. In stratified cross-validation, the original dataset is divided into five mutually exclusive subsets (folds), such that the proportion of classes in each fold matches that of the entire dataset.

For each classifier, we trained four folds and evaluated the remaining fold, iterating this process five times so that each fold serves once as the validation set. This ensures that every sample contributes exactly once to an out-of-sample

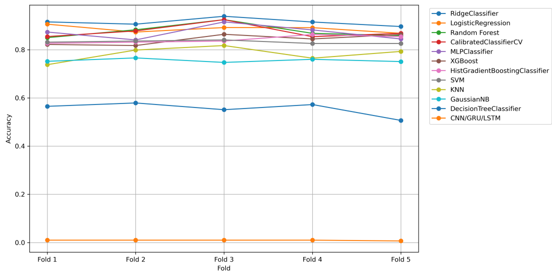

Figure 6: Accuracy across five folds for each classifier.

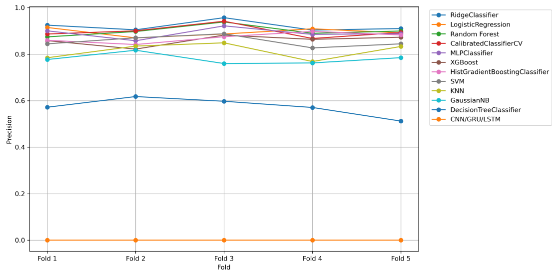

Figure 7: Precision across five folds for each classifier.

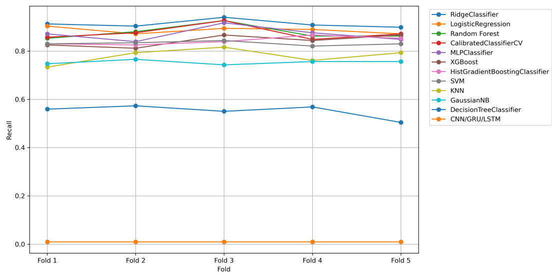

Figure 8: Recall across five folds for each classifier.

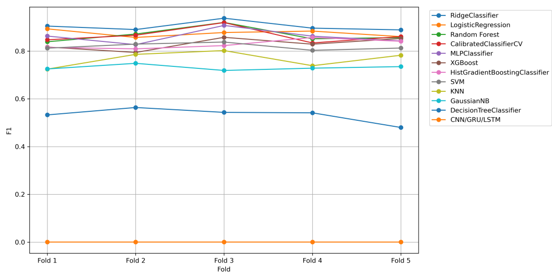

Figure 9: F1-Score across five folds for each classifier.

evaluation and reduces the variance in the performance estimates compared with a single train–test split.

The fold-wise results for each classifier and metric are listed below. For clarity, five-fold values (Fold 1 to 5) are provided for accuracy, precision, recall, and F1Score:

Figure 6 through 9 present the foldwise distributions for each metric across all classifiers. These plots display five-fold values as points (one point per classifier per fold), allowing the visualization of variability across folds.

Table 5 summarizes the cross-validation means and their 95% CIs for accuracy, precision, recall, and F1-Score for all 12 classifiers. Figures 10 through 13 show the same results graphically, with error bars indicating the 95% CI for each classifier's mean metric value.

7.2 One-way analysis of variance (ANOVA)

To determine whether the differences among the classifiers' mean metrics are statistically significant, we conducted one-way ANOVA tests for each performance metric (accuracy, precision, Recall, F1-Score). The null hypothesis for each test is that all 12 classifiers have the same true mean for the given metric, while the alternative hypothesis is that at least one classifier's mean differs.

Table 4: One-way ANOVA across twelve classifiers for each performance metric

Metric	**F-Statistic**	**Degrees of Freedom**	**p-Value**
Accuracy	665.8480	$(11, 48)$	1.6479×10^{-48}
Precision	533.5686	$(11, 48)$	3.2028×10^{-46}
Recall	639.1222	$(11, 48)$	4.3709×10^{-48}
F1-Score	541.7493	$(11, 48)$	2.2307×10^{-46}

ANOVA partitions the total variance of the fold-wise scores into between-group variance (variance of classifier means around the grand mean) and within-group variance (variance of fold-wise scores around the mean of each classifier). The resulting F-statistic and p-value indicate whether the observed differences in means exceeded what would be expected by random chance.

Table 4 summarizes the ANOVA results. For each metric, the F-statistic, degrees of freedom, and p-value are reported. In all cases, the F-statistic is very large, and the p-value is effectively zero, showing that at least one classifier's mean differs significantly from the others.

These results indicate that for every metric (accuracy, precision, Recall, F1-Score), the null hypothesis of equal means across all classifiers can be rejected at p < 0.001. In other words, the differences in the mean performance among the classifiers were highly significant.

7.3 95% confidence intervals

For each classifier and evaluation metric (Accuracy, Precision, Recall, and F1-score), we computed the sample mean $\bar{x}$ and sample standard deviation s across the five cross-validation folds. Given the small sample size ($n = 5$), the variability across folds must be interpreted with care. Instead of assuming a normal distribution, we used the Student's t distribution, which is more appropriate for small samples, to calculate the confidence intervals (CIs). Specifically, the 95% CI for each metric is given by:

$$\bar{x} \pm t_{0.975,\, df=4} \times \frac{s}{\sqrt{5}},$$

where $t_{0.975,\, df=4} \approx 2.776$ is the critical value from the t-distribution with 4 degrees of freedom.

The resulting confidence interval provides a statistical range within which the true mean performance metric is expected to lie with 95% confidence. This is particularly important in classification tasks involving EEG data, where performance can vary significantly across folds due to inter-session and inter-subject variability.

In practical terms, a narrower confidence interval indicates higher consistency and reliability of a model's performance, while a wider interval reflects greater variability and uncertainty. Traditional machine learning models, such as Ridge Regression and Logistic Regression, exhibited relatively narrow confidence intervals across all metrics, suggesting stable and robust performance across folds. In contrast, deep learning models like LSTM and GRU showed wider intervals, indicating higher variability—likely due to the mismatch between their design (which favors raw temporal data) and the input feature format (aggregated statistical features).

Table 5 reports the mean values alongside their corresponding margin of error, computed as $t_{0.975,4}\left(s/\sqrt{5}\right)$. These results offer a more statistically grounded comparison of model performance, supplementing the cross-validation metrics presented earlier.

Confidence intervals also serve as a basis for subsequent statistical testing. In this study, they complement the ANOVA analysis described in Section 5.7 by providing insight into both the central tendency and the variability of each classifier's performance.

7.4 Evaluation results and statistical analysis

All the classifiers were retrained on the full training set before the final evaluation of the held-out test data. Table 6 summarizes the test-set performance in terms of accuracy, precision, recall, and F1-score (all in percentage). The bar plots in Figures 14–17 compare these metrics across classifiers. Based on these results, the classifiers fell into distinct performance tiers.

– **In the top-tier group,** RidgeClassifier achieved the highest overall performance, with the highest accuracy

Table 5: Cross-validation means and 95% confidence intervals for all classifiers

Classifier	Accuracy (95% CI)	Precision (95% CI)	Recall (95% CI)	F1-Score (95% CI)
LogisticRegression	0.8867 ± 0.0191	0.8946 ± 0.0223	0.8863 ± 0.0177	0.8746 ± 0.0190
DecisionTree	0.5552 ± 0.0359	0.5741 ± 0.0492	0.5514 ± 0.0343	0.5323 ± 0.0389
GaussianNB	0.7556 ± 0.0095	0.7803 ± 0.0286	0.7541 ± 0.0111	0.7314 ± 0.0141
Random Forest	0.8773 ± 0.0365	0.9001 ± 0.0295	0.8771 ± 0.0366	0.8678 ± 0.0391
HistGradientBoosting	0.8427 ± 0.0195	0.8716 ± 0.0270	0.8431 ± 0.0211	0.8294 ± 0.0244
MLP	0.8717 ± 0.0379	0.8908 ± 0.0291	0.8706 ± 0.0377	0.8606 ± 0.0376
Ridge	0.9148 ± 0.0196	0.9203 ± 0.0277	0.9129 ± 0.0201	0.9037 ± 0.0250
Calibrated	0.8764 ± 0.0361	0.8981 ± 0.0340	0.8761 ± 0.0376	0.8658 ± 0.0405
XGBoost	0.8427 ± 0.0275	0.8599 ± 0.0287	0.8431 ± 0.0305	0.8303 ± 0.0324
SVM	0.8324 ± 0.0080	0.8550 ± 0.0299	0.8321 ± 0.0103	0.8192 ± 0.0175
KNN	0.7828 ± 0.0387	0.8141 ± 0.0437	0.7798 ± 0.0401	0.7670 ± 0.0414
CNN/GRU/LSTM	0.0092 ± 0.0018	0.0001 ± 0.0000	0.0092 ± 0.0000	0.0002 ± 0.0000

and F1-score on the test set. Logistic Regression and Random Forest also performed well, with similarly high precision and recall. These three models consistently outperformed the others across most metrics, indicating that they captured relevant patterns in the data more effectively. Their superiority is evident in Table 6 and the tall bars for these models in Figures 14–17. The confidence intervals of the top-tier classifiers (Figures 10–13) are relatively narrow, reflecting a stable performance across the cross-validation folds.

– **Mid-tier classifiers**, multilayer perceptron (MLP), CalibratedClassifierCV, XGBoost, and support vector machine (SVM) – achieved moderate performance. Their test accuracies and F1-scores were lower than those of the top-tier models but higher than those of the remaining classifiers. Precision and recall for this group tended to be acceptable but showed more variability (visible in the fold-wise plots in Figures 6–9) than the top group. In Figures 14–17, the mid-tier models produced mid-height bars. Overall, this group indicated a strong performance capability, but with less consistency and slightly lower averages than the leaders.

– **The lower-tier classifiers** include the HistGradientBoostingClassifier, K-Nearest Neighbors (KNN), and Gaussian Naive Bayes. These models achieved the next level of performance, with notably lower accuracy and F1 scores than those of the mid-tier group. Their test set results (Table 6) show substantial drops in one or more metrics. For example, KNN and GaussianNB have lower precision and recall than the top groups, and the bars for these models in Figures 14–17 are noticeably shorter. The confidence intervals (Figures 10–13) for the lower-tier models were wider, indicating greater variability across folds. This suggests that these classifiers are less stable, perhaps because of their sensitivity to data variations or model assumptions.

– **The underperformers** consisted of the Decision Tree Classifier and three deep learning models (CNN, GRU, and LSTM). These models yielded the lowest test performance. The Decision Tree had particularly low scores, and the CNN/GRU/LSTM models failed to match the performance of even the lower-tier traditional classifiers. In the bar plots (Figures 14–17), these models produced the shortest bars, and Table 6 shows their metrics near the bottom. The fold-wise variability plots (Figures 6–9) show large fluctuations for these models and their confidence intervals (Figures 10–13) are the widest, indicating inconsistent performance. In summary, these classifiers were not as generalized to the test set as the others.

Performance stability and statistical significance were assessed across all classifiers. Figures 6–9 show the variability of each metric across cross-validation folds: top-tier models have relatively tight distributions, whereas lower-tier and underperforming models show a wider spread. Figures 10–13 show the 95% confidence intervals for each metric, again highlighting that the best models have smaller error bars. A one-way ANOVA was conducted for each metric to test for significant differences among classifiers. Table 4 reports the F-statistics and p-values. All ANOVA tests were significant ($p < 0.05$), confirming that at least some classifiers differed in terms of accuracy, precision, recall, and F1-score.

Notably, the classifiers that performed best during cross-validation (RidgeClassifier, Logistic Regression, and Random Forest in the top tier) also achieved the highest scores on the held-out test set. This indicates that our cross-validation assessment reliably identified the strongest models, and that these models generalize well to new data. In contrast, the underperforming models remained the lowest in both validation and test metrics, demonstrating that differences in performance observed during training were predictive of the final test performance.

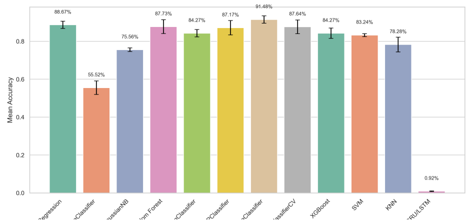

Figure 10: Accuracy per classifier with 95% confidence intervals

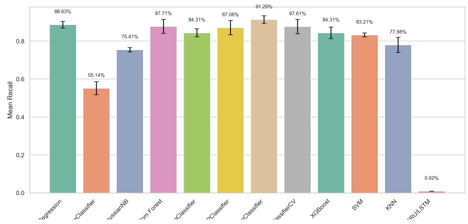

Figure 11: Recall per classifier with 95% confidence intervals

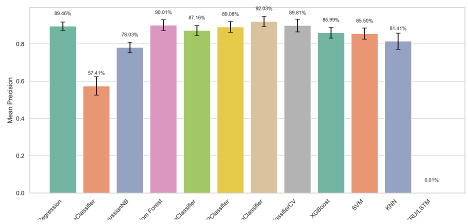

Figure 12: Precision per classifier with 95% confidence intervals

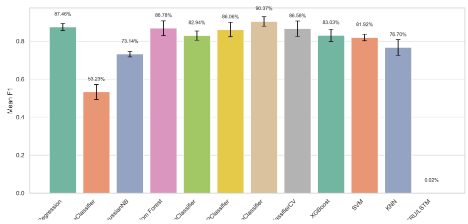

Figure 13: F1 per classifier with 95% confidence intervals

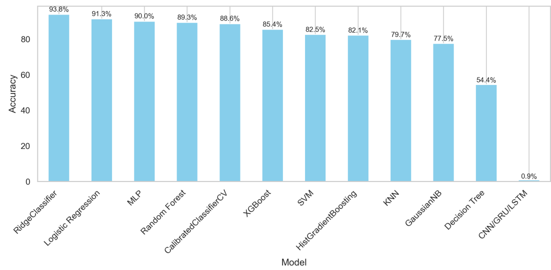

Figure 14: Bar-plot comparison of testing-set accuracy for all classifiers

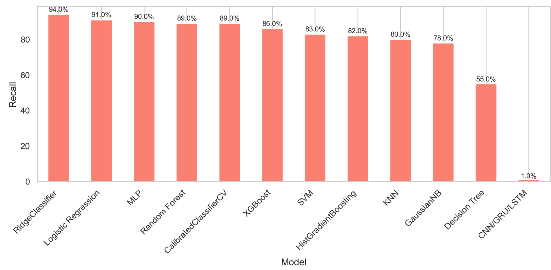

Figure 15: Bar-plot comparison of testing-set recall for all classifiers

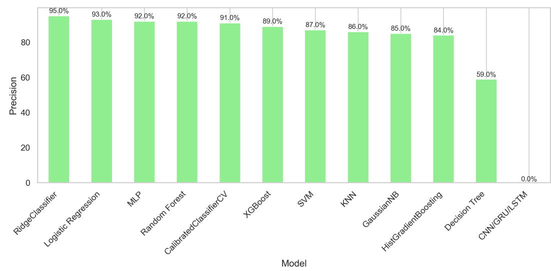

Figure 16: Bar-plot comparison of testing-set precision for all classifiers

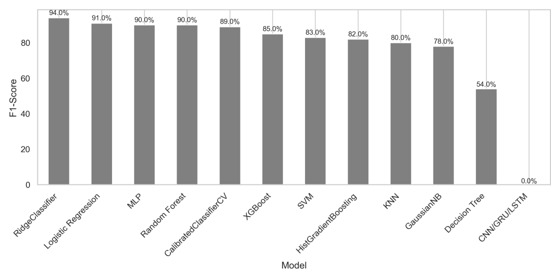

Figure 17: Bar-plot comparison of testing-set F1-score for all classifiers

8 Discussion

A comparative analysis of various classification models for EEG-based biometric authentication reveals significant insights into the interplay between the model architecture, data characteristics, and feature representation. Notably, traditional machine learning models, particularly the Ridge Classifier and Logistic Regression, demonstrated superior performance, achieving accuracies of 93.8% and 91.27%, respectively. This performance surpasses that of more complex deep learning models, which is counterintuitive given the latter's capacity to model intricate patterns.

The efficacy of the ridge classifier and Logistic Regression can be attributed to several factors. First, these linear models are inherently robust to high-dimensional data and are less prone to overfitting, making them well suited for EEG data characterized by high dimensionality and noise. Second, the preprocessing steps involving statistical, spectral, and time-frequency domain transformations likely enhanced the linear separability of the data, aligning well with the assumptions of these models. This suggests that, with appropriate feature engineering, EEG signals can be effectively modeled using linear classifiers.

The multilayer perceptron (MLP), which achieved an accuracy of 89.96%, also performed commendably. As a relatively simple neural network, MLP benefits from its capacity to model nonlinear relationships without the complexity and data requirements of deeper architectures. Its performance indicates that moderately complex models can effectively capture essential patterns in pre-processed EEG data.

Ensemble methods, such as random forest and CalibratedClassifierCV, also exhibited strong performance, underscoring their ability to handle nonlinearities and interactions between features. These models are known for their robustness to noise and overfitting, which are particularly beneficial when dealing with physiological data, such as EEG.

Moderate performance was observed with models, such as XGBoost (85.37%), SVM (82.53%), and HistGradientBoosting (82.1%). Although these models are adept at capturing complex patterns, their performance may be hindered by the nature of EEG data, which can be noisy and exhibit complex interdependencies that are challenging to model without extensive data and careful tuning.

Instance-based and probabilistic models, such as KNN (79.69%) and GaussianNB (77.51%), underperformed, likely because of their underlying assumptions. KNN's reliance of KNN on distance metrics can be problematic in high-dimensional spaces, and GaussianNB's assumption of feature independence and normality rarely holds in EEG data, which often exhibit inter-feature dependencies and non-Gaussian distributions.

Surprisingly, deep learning models—CNN, LSTM, GRU, and Bidirectional LSTM—achieved the lowest accuracies, hovering around 87%. Although these models are renowned for their representation-learning capabilities, their poor performance in this study stems from fundamental design constraints:

– The deep learning models were trained on extracted statistical features rather than raw EEG time-series data. This limits their ability to learn temporal or spatial dependencies, which are core strengths of architectures such as LSTM and CNN.

– The dataset size was relatively small for training deep learning models effectively, particularly when using high-capacity architectures without temporal input sequences.

It is important to emphasize that applying deep models to pre-aggregated statistical features restricts their potential to exploit temporal and sequential information inherent in raw EEG signals. This design decision creates a fundamental mismatch between the model architecture and the nature of the input data, making the lower performance of deep models unsurprising. Therefore, the results should not be interpreted as a definitive comparison between traditional machine learning and deep learning models, but rather as an evaluation of these methods under a constrained and non-ideal setup for deep architectures.

These findings align with the literature, which shows that deep learning models can achieve high accuracy when trained on raw EEG data and with sufficient volume. For instance, studies have demonstrated that with as little as 1–3 seconds of raw EEG data per participant, models can achieve over 95% accuracy, provided that appropriate signal structures are preserved and learned [58]. This highlights the importance of both data format and quantity in evaluating deep learning effectiveness.

Moreover, while statistical features may suffice for traditional classifiers, they do not retain the rich temporal dynamics that deep models are specifically designed to capture. Feeding deep networks with flattened, aggregated input features inherently undermines their advantage in learning complex time-dependent patterns.

Another contributing factor is the sensitivity of deep models to architectural choices and hyperparameters. Suboptimal tuning can further degrade performance, especially when combined with non-temporal input and limited data. Additionally, the computational demands of training deep networks can make them less practical in small-scale studies or low-resource settings.

In contrast, traditional machine-learning models offer strengths in interpretability, efficiency, and robustness under limited data conditions. Their superior performance in this study reflects a better match between their design and the structure of the extracted features.

Looking forward, future research should aim to apply deep learning models to raw EEG time-series data, where their full potential in capturing temporal dependencies can be evaluated. Approaches such as data augmentation, transfer learning, and task-specific architectures may also help address limitations related to data size and diversity.

In summary, the success of a classifier in EEG-based authentication is closely linked to the alignment between data characteristics and model architecture. Traditional models perform effectively on statistical features, whereas deep models require raw, sequential data and larger datasets to demonstrate their advantages. These insights will guide future work in designing better-suited pipelines for neurobiometric systems.

9 Conclusion

The study assessed EEG-based biometric authentication through a complete classification system evaluation. Multiple machine learning and deep learning models were evaluated using a structured pipeline that included data preprocessing, feature extraction, and classifier evaluation. The results showed that traditional machine learning models,

Table 6: Performance on held-out testing set for all twelve classifiers

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RidgeClassifier	93.80	95	94	94
Logistic Regression	91.27	93	91	91
MLPClassifier	89.96	92	90	90
Random Forest	89.30	92	89	90
CalibratedClassifierCV	88.60	91	89	89
XGBoost	85.37	89	86	85
SVM	82.53	87	83	83
HistGradientBoosting	82.10	84	82	82
KNN	79.69	86	80	80
GaussianNB	77.51	85	78	78
DecisionTreeClassifier	54.37	59	55	54
CNN/GRU/LSTM	0.87	0	1	0

including ridge classifiers, logistic regression, and MLP, achieved better accuracy and reliability than deep learning approaches. The ensemble methods, Random Forest and Calibrated Classifier CV demonstrated strong performance because they excel at detecting complex EEG patterns. The deep learning models CNN, LSTM, and GRU achieved significantly lower accuracy because feature extraction proved difficult and the dataset size remained limited.

Research indicates that EEG-based authentication works best through well-optimized machine learning models that create a secure and reliable biometric verification system. Future research should investigate the development of hybrid models that combine feature engineering techniques with deep learning methods to boost the classification accuracy. The performance of deep learning models in EEG biometrics can be enhanced using larger dataset sizes and sophisticated augmentation methods. This research adds to the EEG authentication literature while offering essential guidance for selecting appropriate models for real-world implementations.

References

[1] P. Bontrager, A. Roy, J. Togelius, N. Memon, and A. Ross, “Deepmasterprints: Generating masterprints for dictionary attacks via latent variable evolution,” in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2018, pp. 1–9. [Online]. Available: https://doi.org/10.1109/btas.2018.8698539

[2] N. K. Ratha, J. H. Connell, and R. M. Bolle, “Enhancing security and privacy in biometrics-based authentication systems,” *IBM systems Journal*, vol. 40, no. 3, pp. 614–634, 2001. [Online]. Available: https://doi.org/10.1147/sj.403.0614

[3] S. Marcel and J. d. R. Millán, “Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 4, pp. 743–752, 2007. [Online]. Available: https://doi.org/10.1109/TPAMI.2007.1013

[4] M. DelPozo-Banos, C. M. Travieso, C. T. Weidemann, and J. B. Alonso, “Eeg biometric identification: a thorough exploration of the time-frequency domain,” *Journal of neural engineering*, vol. 12, no. 5, p. 056019, 2015. [Online]. Available: https://doi.org/10.1088/1741-2560/12/5/056019

[5] G.-C. Yang, “Next-generation personal authentication scheme based on eeg signal and deep learning,” *Journal of Information Processing Systems*, vol. 16, no. 5, pp. 1034–1047, 2020.

[6] S. Zhang, L. Sun, X. Mao, C. Hu, and P. Liu, “Review on eeg-based authentication technology,” *Computational intelligence and neuroscience*, vol. 2021, no. 1, p. 5229576, 2021. [Online]. Available: https://doi.org/10.1155/2021/5229576

[7] K. P. Thomas and A. P. Vinod, “Eeg-based biometric authentication using gamma band power during rest state,” *Circuits, Systems, and Signal Processing*, vol. 37, no. 1, pp. 277–289, 2018. [Online]. Available: https://doi.org/10.1007/s00034-017-0551-4

[8] S. Amlani, S. Jaiswal, and S. Patil, “Session authentication using color scheme,” *Proc. Int. J. Comput. Sci. Inf. Technol.(IJCSIT)*, vol. 6, no. 2, pp. 1420–1423, 2015.

[9] C. Priya, “Behavioral biometrics based authentication system using mlp-nn and mvpa,” in *2021 IEEE International Power and Renewable Energy Conference (IPRECON)*. IEEE, 2021, pp. 1–6. [Online]. Available: https://doi.org/10.1109/iprecon52453.2021.9640941

[10] Y. S. Soni, S. Somani, and V. Shete, “Biometric user authentication using brain waves,” in *2016 International Conference on Inventive Computation Technologies (ICICT)*, vol. 2. IEEE, 2016, pp. 1–6. [Online]. Available: https://doi.org/10.1109/inventive.2016.7824888

[11] K. Lis, E. Niewiadomska-Szynkiewicz, and K. Dziewulska, “Siamese neural network for keystroke dynamics-based authentication on partial passwords,” *Sensors*, vol. 23, no. 15, p. 6685, 2023. [Online]. Available: https://doi.org/10.3390/s23156685

[12] M. M. Pandi and A. Valarmathi, “A secured graphical password authentication system,” *International Journal of Engineering Research & Technology (IJERT)*, vol. 2, no. 5, pp. 1013–1019, 2013.

[13] H. Om and M. R. Reddy, “Geometric based remote password authentication using biometrics,” *Journal of Discrete Mathematical Sciences and Cryptography*,

vol. 16, no. 4-5, pp. 207–220, 2013. [Online]. Available: https://doi.org/10.1080/09720529.2013.778459

[14] W. Khalifa, A. Salem, M. Roushdy, and K. Revett, "A survey of eeg based user authentication schemes," in *2012 8th International Conference on Informatics and Systems (INFOS)*. IEEE, 2012, pp. BIO–55.

[15] Z. Ni, A. C. Yuksel, X. Ni, M. I. Mandel, and L. Xie, "Confused or not confused? disentangling brain activity from eeg data using bidirectional lstm recurrent neural networks," in *Proceedings of the 8th acm international conference on bioinformatics, computational biology, and health informatics*, 2017, pp. 241–246.

[16] P. Wang and J. Hu, "A hybrid model for eeg-based gender recognition," *Cognitive neurodynamics*, vol. 13, no. 6, pp. 541–554, 2019. [Online]. Available: https://doi.org/10.1007/s11571-019-09543-y

[17] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using eeg signals: A survey," *IEEE transactions on affective computing*, vol. 10, no. 3, pp. 374–393, 2017. [Online]. Available: https://doi.org/10.1109/taffc.2017.2714671

[18] Y. Höller and A. Uhl, "Do eeg-biometric templates threaten user privacy?" in *Proceedings of the 6th ACM workshop on information hiding and multimedia security*, 2018, pp. 31–42. [Online]. Available: https://doi.org/10.1145/3206004.3206006

[19] S. Mueller, D. Wang, M. D. Fox, B. T. Yeo, J. Sepulcre, M. R. Sabuncu, R. Shafee, J. Lu, and H. Liu, "Individual variability in functional connectivity architecture of the human brain," *Neuron*, vol. 77, no. 3, pp. 586–595, 2013. [Online]. Available: https://doi.org/10.1016/j.neuron.2012.12.028

[20] M. Wang, J. Hu, and H. A. Abbass, "Brainprint: Eeg biometric identification based on analyzing brain connectivity graphs," *Pattern Recognition*, vol. 105, p. 107381, 2020. [Online]. Available: https://doi.org/10.1016/j.patcog.2020.107381

[21] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE transactions on image processing*, vol. 23, no. 2, pp. 710–724, 2013. [Online]. Available: https://doi.org/10.1109/tip.2013.2292332

[22] J. Bonneau, C. Herley, P. C. Van Oorschot, and F. Stajano, "The quest to replace passwords: A framework for comparative evaluation of web authentication schemes," in *2012 IEEE symposium on security and privacy*. IEEE, 2012, pp. 553–567. [Online]. Available: https://doi.org/10.1109/sp.2012.44

[23] J. J. Bengson, T. A. Kelley, X. Zhang, J.-L. Wang, and G. R. Mangun, "Spontaneous neural fluctuations predict decisions to attend," *Journal of Cognitive Neuroscience*, vol. 26, no. 11, pp. 2578–2584, 2014.

[24] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "Bci2000: a general-purpose brain-computer interface (bci) system," *IEEE Transactions on biomedical engineering*, vol. 51, no. 6, pp. 1034–1043, 2004. [Online]. Available: https://doi.org/10.1109/tbme.2004.827072

[25] M. Ruiz-Blondet, N. Khlaifian, B. Armstrong, Z. Jin, K. Kurtz, and S. Laszlo, "Brainprint: Identifying unique features of neural activity with machine learning," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 36, no. 36, 2014.

[26] X. Jin, J. Tang, X. Kong, Y. Peng, J. Cao, Q. Zhao, and W. Kong, "Ctnn: A convolutional tensor-train neural network for multi-task brainprint recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 103–112, 2020. [Online]. Available: https://doi.org/10.1109/tnsre.2020.3035786

[27] S. B. Salem and Z. Lachiri, "Cnn-svm approach for eeg-based person identification using emotional dataset," in *2019 International Conference on Signal, Control and Communication (SCC)*. IEEE, 2019, pp. 241–245. [Online]. Available: https://doi.org/10.1109/scc47175.2019.9116175

[28] S. Altahat, G. Chetty, D. Tran, and W. Ma, "Analysing the robust eeg channel set for person authentication," in *International conference on neural information processing*. Springer, 2015, pp. 162–173.

[29] Q. Wu, Y. Zeng, C. Zhang, L. Tong, and B. Yan, "An eeg-based person authentication system with open-set capability combining eye blinking signals," *Sensors*, vol. 18, no. 2, p. 335, 2018. [Online]. Available: https://doi.org/10.3390/s18020335

[30] X. Zhang, L. Yao, C. Huang, T. Gu, Z. Yang, and Y. Liu, "Deepkey: A multimodal biometric authentication system via deep decoding gaits and brainwaves," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, no. 4, pp. 1–24, 2020.

[31] J. Li, Z. Zhang, and H. He, "Implementation of eeg emotion recognition system based on hierarchical convolutional neural networks," in *Advances in Brain Inspired Cognitive Systems: 8th International Conference, BICS 2016, Beijing, China, November 28-30, 2016, Proceedings 8*. Springer, 2016, pp. 22–33.

[32] G. Rodriguez-Bermudez and P. J. Garcia-Laencina, "Analysis of eeg signals using nonlinear dynamics and chaos: a review," *Applied mathematics & information sciences*, vol. 9, no. 5, p. 2309, 2015.

[33] P. Kumar, R. Saini, B. Kaur, P. P. Roy, and E. Scheme, "Fusion of neuro-signals and dynamic signatures for person authentication," *Sensors*, vol. 19, no. 21, p. 4641, 2019. [Online]. Available: https://doi.org/10.3390/s19214641

[34] D. Nguyen, D. Tran, D. Sharma, and W. Ma, "On the study of eeg-based cryptographic key generation," *Procedia computer science*, vol. 112, pp. 936–945, 2017. [Online]. Available: https://doi.org/10.1016/j.procs.2017.08.126

[35] C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein, "Low-cost electroencephalogram (eeg) based authentication," in *2011 5th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2011, pp. 442–445. [Online]. Available: https://doi.org/10.1109/ner.2011.5910581

[36] E. Gupta, M. Agarwal, and R. Sivakumar, "Blink to get in: Biometric authentication for mobile devices using eeg signals," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6. [Online]. Available: https://doi.org/10.1109/icc40277.2020.9148741

[37] S. R. K. Gopal and D. Shukla, "Concealable biometric-based continuous user authentication system an eeg induced deep learning model," in *2021 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2021, pp. 1–8. [Online]. Available: https://doi.org/10.1109/ijcb52358.2021.9484345

[38] A. Mishra, N. Raj, and G. Bajwa, "Eeg-based image feature extraction for visual classification using deep learning," in *2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*. IEEE, 2022, pp. 181–188. [Online]. Available: https://doi.org/10.1109/idsta55301.2022.9923087

[39] J. Y. Cheng, H. Goh, K. Dogrusoz, O. Tuzel, and E. Azemi, "Subject-aware contrastive learning for biosignals," *arXiv preprint arXiv:2007.04871*, 2020.

[40] B. Bandana Das, S. Kumar Ram, K. Sathya Babu, R. K. Mohapatra, and S. P. Mohanty, "Person identification using autoencoder-cnn approach with multitask-based eeg biometric," *Multimedia Tools and Applications*, vol. 83, no. 35, pp. 83 205–83 225, 2024. [Online]. Available: https://doi.org/10.1007/s11042-024-18693-z

[41] N. G. Venkataswamy and M. H. Imtiaz, "Support vector machine for person classification using the eeg signals," *arXiv preprint arXiv:2411.17446*, 2024. [Online]. Available: https://doi.org/10.1109/icecet58911.2023.10389511

[42] P. Lakhan, N. Banluesombatkul, N. Sricom, K. Surapat, R. Rotruchiphong, P. Sawangjai, T. Yagi, T. Limpiti, and T. Wilaiprasitporn, "Eeg-bbnet: a hybrid framework for brain biometric using graph connectivity," *arXiv preprint arXiv:2208.08901*, 2022. [Online]. Available: https://doi.org/10.1109/lsens.2024.3522981

[43] D. Cherifi, N. Falkoun, F. Ouakouak, L. Boubchir, and A. Nait-Ali, "Eeg signal feature extraction and classification for epilepsy detection," *Informatica*, vol. 46, no. 4, 2022. [Online]. Available: https://doi.org/10.31449/inf.v46i4.3768

[44] R. Sahu, S. R. Dash, and A. Baral, "Identification of students' confusion in classes from eeg signals using convolution neural network," *Informatica*, vol. 48, no. 1, 2024. [Online]. Available: https://doi.org/10.31449/inf.v48i1.4604

[45] P. Campisi and D. La Rocca, "Brain waves for automatic biometric-based user recognition," *IEEE transactions on information forensics and security*, vol. 9, no. 5, pp. 782–800, 2014. [Online]. Available: https://doi.org/10.1109/tifs.2014.2308640

[46] M. L. Martini, E. K. Oermann, N. L. Opie, F. Panov, T. Oxley, and K. Yaeger, "Sensor modalities for brain-computer interface technology: a comprehensive literature review," *Neurosurgery*, vol. 86, no. 2, pp. E108–E117, 2020. [Online]. Available: https://doi.org/10.1093/neuros/nyz286

[47] R. Palaniappan and D. P. Mandic, "Biometrics from brain electrical activity: A machine learning approach," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 4, pp. 738–742, 2007. [Online]. Available: https://doi.org/10.1109/tpami.2007.1013

[48] J. Malmivuo, "Comparison of the properties of eeg and meg in detecting the electric activity of the brain," *Brain topography*, vol. 25, pp. 1–19, 2012. [Online]. Available: https://doi.org/10.1007/s10548-011-0202-1

[49] R. Das, E. Maiorana, and P. Campisi, "Motor imagery for eeg biometrics using convolutional neural network," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2018, pp. 2062–2066. [Online]. Available: https://doi.org/10.1109/icassp.2018.8461909

[50] T. Koike-Akino, R. Mahajan, T. K. Marks, Y. Wang, S. Watanabe, O. Tuzel, and P. Orlik, "High-accuracy user identification using eeg biometrics," in *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2016, pp. 854–858. [Online]. Available: https://doi.org/10.1109/embc.2016.7590835

[51] H. Huang, L. Hu, F. Xiao, A. Du, N. Ye, and F. He, "An eeg-based identity authentication system with audiovisual paradigm in iot," *Sensors*, vol. 19, no. 7, p. 1664, 2019. [Online]. Available: https://doi.org/10.3390/s19071664

[52] S. Keshishzadeh, A. Fallah, and S. Rashidi, "Improved eeg based human authentication system on large dataset," in *2016 24th Iranian Conference on Electrical Engineering (ICEE)*. IEEE, 2016, pp. 1165–1169. [Online]. Available: https://doi.org/10.1109/iraniancee.2016.7585697

[53] Q. Gui, M. V. Ruiz-Blondet, S. Laszlo, and Z. Jin, "A survey on brain biometrics," *ACM Computing Surveys (CSUR)*, vol. 51, no. 6, pp. 1–38, 2019. [Online]. Available: https://doi.org/10.1145/3230632

[54] K. H. Ali, "Motor imagery detection in ecg signals using wavelet packet decomposition and multiscale convolutional neural networks," *Informatica*, vol. 49, no. 12, 2025. [Online]. Available: https://doi.org/10.31449/inf.v49i12.6690

[55] A. Subasi, "Decision support system for epileptic seizure detection using discrete wavelet transform and machine learning algorithms," *Biomedical Engineering Online*, vol. 9, no. 1, pp. 1–17, 2010.

[56] A. K. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005. [Online]. Available: https://doi.org/10.1016/j.patcog.2005.01.012

[57] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[58] C. Gómez-Tapia, B. Bozic, and L. Longo, "On the minimal amount of eeg data required for learning distinctive human features for task-dependent biometric applications," *Frontiers in neuroinformatics*, vol. 16, p. 844667, 2022. [Online]. Available: https://doi.org/10.3389/fninf.2022.844667